混合整数二次錐計画法を用いた回帰式の変数選択 (最適化の基礎理論と応用)

宮代, 隆平; 高野, 祐一

数理解析研究所講究録
第 1879 巻 2014 年 222-229

混合整数二次錐計画法を用いた回帰式の変数選択

宮代 隆平
東京農工大学・大学院工学研究院
Ryuhei Miyashiro
Institute of Engineering,
Tokyo University of Agriculture and Technology

高野 祐一
東京工業大学・大学院社会理工学研究科
Yuichi Takano
Graduate School of Decision Science and Technology,
Tokyo Institute of Technology

1 回帰分析における変数選択と情報量規準

統計学における回帰分析では，変数選択は重要な課題として知られている．本稿では，赤池情報量規準を始めとするいくつかの代表的な情報量規準を考え，それらを最小にする変数選択問題が混合整数二次錐計画問題として定式化できることを示す．

n 個の観測値 $(y_i; x_{i1}, x_{i2}, \ldots, x_{ik})$ $(i = 1, 2, \ldots, n)$ が与えられ，y_i を被説明変数，x_{ij} $(j = 1, 2, \ldots, k)$ を k 種類の説明変数とする．重回帰分析では，以下の線形モデルによって被説明変数の値を予測する：

$$y_i = b + a_1 x_{i1} + a_2 x_{i2} + \cdots + a_k x_{ik} + \varepsilon_i, \tag{1}$$

ただし ε_i は i 番目の観測値の予測残差．

このとき，候補となる p 種類の変数の中から最適な説明変数集合を選択する変数選択問題を考える．

一般に，説明変数を多く含むモデルは与えられたデータに対しては当てはまりが良いものの，未知のデータに対しては予測能力が低くなる．モデルの複雑さ（説明変数の個数）とモデルの当てはまりの良さ（所与のデータに対する誤差の小ささ）とのバランスを考慮した適合度指標の代表的なものとして，赤池情報量規準 (AIC; [1]) がある．回帰モデル (1) に対するいくつかの自然な仮定の

下で，定数項を削除した AIC 値は

$$n \log \left(\frac{1}{n} \sum_{i=1}^{n} \varepsilon_i^2 \right) + 2k \tag{2}$$

と表わされ，この値が小さいほど良いモデルとされる．以下では AIC の最小化を例にとり説明を行うが，AIC 以外にもベイジアン情報量規準 (BIC; [13]) を始め，多数の適合度指標がこれまでに提案されている（[9] などを参照のこと）．

2 AIC 最小化問題

AIC を適合度指標とした説明変数の選択には，ステップワイズ変数選択がしばしば用いられる．ステップワイズ変数選択は短時間で良いモデル（AIC が小さいモデル）を導くことができるが，ヒューリスティクスであるため必ずしも AIC が最小のモデルを選択するわけではない．

一方で k を定数と見なすと，関数 (2) の最小化は $\sum_{i=1}^{n} \varepsilon_i^2$ の最小化と等価になり，これは「p 個の説明変数候補から，残差二乗和を最小化するように k 個の説明変数を選ぶ問題」になる．この問題は以下の混合整数二次計画問題として定式化できることが知られている（例えば [6]）：

$$\begin{aligned}
&\underset{\boldsymbol{a}, b, \boldsymbol{\varepsilon}, \boldsymbol{z}}{\text{minimize}} \quad \sum_{i=1}^{n} \varepsilon_i^2 \\
&\text{subject to} \\
&\varepsilon_i = y_i - \left(b + \sum_{j=1}^{p} a_j x_{ij} \right) \quad (i = 1, 2, \ldots, n), && (3) \\
&-Mz_j \le a_j \le Mz_j \quad (j = 1, 2, \ldots, p), && (4) \\
&\sum_{j=1}^{p} z_j = k, && (5) \\
&z_j \in \{0, 1\} \quad (j = 1, 2, \ldots, p), && (6)
\end{aligned}$$

ただし M は十分に大きい正の定数．

しかしながら，上記の定式化では k を所与の定数として扱っており，現実には AIC を最小にする説明変数の個数 k は事前にはわからないという問題点がある．

3 AIC 最小化の混合整数二次錐計画問題としての定式化

本研究では，選択される説明変数の個数 k が未知の場合の AIC 最小化問題を，混合整数二次錐計画問題 (Mixed Integer SOCP; MISOCP) として定式化する．

AIC 最小化問題は，関数 (2) を最小化すればよいことから，直接的には以下の非線形混合整数計

画問題として定式化できる（k も変数である点に注意）：

$$\begin{aligned} \underset{\boldsymbol{a},\, b,\, \boldsymbol{\varepsilon},\, k,\, \boldsymbol{z}}{\text{minimize}} \quad & n \log \left(\frac{1}{n} \sum_{i=1}^{n} \varepsilon_i^2 \right) + 2k \\ \text{subject to} \quad & \text{制約式 (3), (4), (5), (6).} \end{aligned} \tag{7}$$

しかし，この定式化の目的関数 (7) は非凸な非線形関数であり，0-1 変数 z_j の整数性を緩和した上でもこの問題を解くことは困難である．整数計画問題に対して分枝限定法がうまく働くには，その連続緩和問題が効率的に解ける形になっている必要がある．以下ではそれを念頭におき，問題を等価変形する．

目的関数 (7) を以下のように変換する．

$$\begin{aligned} & \underset{\boldsymbol{a},\, b,\, \boldsymbol{\varepsilon},\, k,\, \boldsymbol{z}}{\text{minimize}} \quad n \log \left(\frac{1}{n} \sum_{i=1}^{n} \varepsilon_i^2 \right) + 2k \\ \Longleftrightarrow \quad & \underset{\boldsymbol{a},\, b,\, \boldsymbol{\varepsilon},\, k,\, \boldsymbol{z}}{\text{minimize}} \quad \exp \left(\log \left(\frac{1}{n} \sum_{i=1}^{n} \varepsilon_i^2 \right) + \frac{2k}{n} \right) \\ \Longleftrightarrow \quad & \underset{\boldsymbol{a},\, b,\, \boldsymbol{\varepsilon},\, k,\, \boldsymbol{z}}{\text{minimize}} \quad \exp \left(\frac{2k}{n} \right) \cdot \sum_{i=1}^{n} \varepsilon_i^2 \end{aligned}$$

項 $\exp(2k/n) \cdot \sum_{i=1}^{n} \varepsilon_i^2$ の上界を表す連続変数 f を導入することにより，問題は以下の形に変形できる：

$$\begin{aligned} \underset{\boldsymbol{a},\, b,\, \boldsymbol{\varepsilon},\, f,\, k,\, \boldsymbol{z}}{\text{minimize}} \quad & f \\ \text{subject to} \quad & \sum_{i=1}^{n} \varepsilon_i^2 \le f \cdot \exp \left(-\frac{2k}{n} \right), \\ & \text{制約式 (3), (4), (5), (6).} \end{aligned} \tag{8}$$

次に，制約式 (8) の非線形性を解消する．新たな 0-1 変数 w_j $(j = 0, 1, \ldots, p)$ を定義し，w_j に関して制約式 $\sum_{j=0}^{p} (w_j \cdot j) = k$ および $\sum_{j=0}^{p} w_j = 1$ を付け加える．すると，制約式 (5) と (6) より k は常に整数値を取るため，w_j は「$j = k$ の時，またその時のみ $w_j = 1$」を満たす 0-1 変数となる．この w_j を用いることにより，非線形の制約式 (8) は以下の線形制約式で表現できる（この線形化手法は，整数計画の分野では SOS type 1 [4, 5] として知られている）：

$$\begin{aligned} & \sum_{i=1}^{n} \varepsilon_i^2 \le f \cdot \exp \left(-\frac{2k}{n} \right) \\ \Longleftrightarrow \quad & \sum_{i=1}^{n} \varepsilon_i^2 \le f \cdot g, \quad g = \exp \left(-\frac{2k}{n} \right) \\ \Longleftrightarrow \quad & \begin{cases} \displaystyle \sum_{i=1}^{n} \varepsilon_i^2 \le f \cdot g, \quad g = \sum_{j=0}^{p} \left(w_j \cdot \exp \left(-\frac{2j}{n} \right) \right), \\ \displaystyle k = \sum_{j=0}^{p} (w_j \cdot j), \quad \sum_{j=0}^{p} w_j = 1. \end{cases} \end{aligned}$$

ここで $g=\sum_{j=0}^{p}(w_j\cdot\exp(-2j/n))$ は，変数 w_j に関して線形であることに留意されたい．制約式 $\sum_{i=1}^{n}\varepsilon_i^2\le f\cdot g$ は非線形であるが，これは以下に示す通り二次錐制約として表現できることが知られている [10]（ただしここで $\|\boldsymbol{u}\|=(\boldsymbol{u}^\top\boldsymbol{u})^{1/2}$）：

$$
\begin{aligned}
&\boldsymbol{\varepsilon}^\top\boldsymbol{\varepsilon}\le f\cdot g,\ f\ge 0,\ g\ge 0\\
\iff\quad &\left\|\begin{pmatrix}2\boldsymbol{\varepsilon}\\ f-g\end{pmatrix}\right\|\le f+g.
\end{aligned}
$$

これらをまとめ，AIC 最小化問題の定式化として以下を得る：

$$
\begin{aligned}
&\underset{\substack{\boldsymbol{a},b,\boldsymbol{\varepsilon},f\\ g,k,\boldsymbol{w},\boldsymbol{z}}}{\text{minimize}}\quad f\\
&\text{subject to}\\
&\varepsilon_i=y_i-\left(b+\sum_{j=1}^{p}a_jx_{ij}\right)\quad(i=1,2,\ldots,n),\\
&\sum_{i=1}^{n}\varepsilon_i^2\le f\cdot g,\quad g=\sum_{j=0}^{p}\left(w_j\cdot\exp\left(-\frac{2j}{n}\right)\right),\\
&\sum_{j=0}^{p}(w_j\cdot j)=k,\quad \sum_{j=0}^{p}w_j=1,\quad \sum_{j=1}^{p}z_j=k,\\
&-Mz_j\le a_j\le Mz_j\quad(j=1,2,\ldots,p),\\
&w_j\in\{0,1\}\quad(j=0,1,\ldots,p),\\
&z_j\in\{0,1\}\quad(j=1,2,\ldots,p).
\end{aligned}
$$

この定式化で表される問題は，線形制約，二次錐制約，整数制約の下で線形目的関数を最小化する MISOCP である．MISOCP の連続緩和問題は二次錐計画問題 (SOCP; [2]) であり，内点法を用いることにより多項式時間で解を求めることができる．また，内点法をベースにした分枝限定法が実装された，MISOCP を扱うことのできる数理計画ソルバーが複数存在する．

なお上記の定式化のうち，制約式 $g=\sum_{j=0}^{p}(w_j\cdot\exp(-2j/n))$ を $g=\sum_{j=0}^{p}(w_j\cdot\exp((-j\log n)/n))=\sum_{j=0}^{p}(w_j\cdot n^{-j/n})$ に置きかえることにより，BIC 最小化問題を MISOCP として定式化できる．また，AIC の有限修正 (corrected AIC; [14])，Hannan-Quinn 情報量規準 (HQ; [7])，residual 情報量規準 (RIC) などの様々な情報量規準の最小化問題，あるいは自由度調整済決定係数 $\bar{R}^2$ [15] の最大化問題も同様に MISOCP として定式化できるが，ここでは省略する．詳細は [12] を参照されたい．

4 計算機実験および考察

計算機実験では，AIC の最小化問題について，ステップワイズ変数選択との比較を通じて提案手法の有効性を検証した．実験には UCI Machine Learning Repository [3] からの 10 つのデータを使用した (表 1).

表1　データのリスト

略称	n	p	データセット [3]
`Housing`	506	13	Housing
`Servo`	167	19	Servo
`AutoMPG`	392	25	Auto MPG
`SolarFlareC`	1066	26	Solar Flare (C-class flares production)
`SolarFlareM`	1066	26	Solar Flare (M-class flares production)
`SolarFlareX`	1066	26	Solar Flare (X-class flares production)
`BreastCancer`	194	32	Breast Cancer Wisconsin
`ForestFires`	517	63	Forest Fires
`Automobile`	159	65	Automobile
`Crime`	1993	100	Communities and Crime

AIC の最小化に際して，MISOCP の求解には数理計画ソルバー CPLEX 12.5 [8] を使用した．計算機環境は，CPU: Intel Xeon W5590 3.33 GHz×2; RAM: 24 GB; OS: 64bit Windows 7 Ultimate SP1; chipset: Intel 5520 のマシンを使用し，各 MISOCP に対して並列分枝限定法 (8 スレッド) を最大 10000 秒実行した．また比較のため，MATLAB R2012b [11] の `LinearModel.stepwise` 関数を用いて，ステップワイズ選択による変数選択を行った．こちらの計算機環境は CPU: Intel Core i7-2600S 2.80 GHz; RAM: 8 GB; OS: 64bit Windows 7 Professional SP1; chipset: Intel Q67 Express である．

表 2　AIC 最小化の実験結果

データ	n	p	手法	AIC 値	k	計算時間 (s)
Housing	506	13	$\mathrm{SWR_{const}}$	**776.36**	11	1.31
			$\mathrm{SWR_{all}}$	**776.36**	11	0.51
			MISOCP	**776.36**	11	10.62
Servo	167	19	$\mathrm{SWR_{const}}$	**258.66**	9	1.85
			$\mathrm{SWR_{all}}$	266.36	14	0.75
			MISOCP	**258.66**	9	8.41
AutoMPG	392	25	$\mathrm{SWR_{const}}$	**333.22**	15	3.96
			$\mathrm{SWR_{all}}$	339.44	19	1.59
			MISOCP	**333.22**	15	51.23
SolarFlareC	1066	26	$\mathrm{SWR_{const}}$	**2816.34**	9	3.09
			$\mathrm{SWR_{all}}$	2819.73	13	4.02
			MISOCP	**2816.34**	9	227.25
SolarFlareM	1066	26	$\mathrm{SWR_{const}}$	**2926.93**	7	2.38
			$\mathrm{SWR_{all}}$	**2926.93**	7	5.99
			MISOCP	**2926.93**	7	92.18
SolarFlareX	1066	26	$\mathrm{SWR_{const}}$	**2882.81**	3	1.20
			$\mathrm{SWR_{all}}$	**2882.81**	3	7.59
			MISOCP	**2882.81**	3	10.73
BreastCancer	194	32	$\mathrm{SWR_{const}}$	509.72	8	3.07
			$\mathrm{SWR_{all}}$	510.58	14	8.13
			MISOCP	**508.73**	10	> 10000
ForestFires	517	63	$\mathrm{SWR_{const}}$	**1429.81**	12	9.56
			$\mathrm{SWR_{all}}$	**1429.81**	12	71.19
			MISOCP	1430.25	13	> 10000
Automobile	159	65	$\mathrm{SWR_{const}}$	−26.87	21	14.20
			$\mathrm{SWR_{all}}$	−47.50	38	24.75
			MISOCP	**−58.49**	32	> 10000
Crime	1993	100	$\mathrm{SWR_{const}}$	3424.26	41	84.94
			$\mathrm{SWR_{all}}$	**3410.92**	50	312.82
			MISOCP	3419.65	51	> 10000

表 2 に，AIC 最小化の計算機実験の結果を示す．「手法」はそれぞれ

- $\mathrm{SWR}_{\mathrm{const}}$: 初期変数集合が切片のみのステップワイズ変数選択
- $\mathrm{SWR}_{\mathrm{all}}$: 初期変数集合が全変数からのステップワイズ変数選択
- MISOCP: MISOCP を用いた提案手法

を示し，AIC 値が最もよかった手法の値を太字で示してある．また k は選択された説明変数の個数である．計算時間について，MISOCP の計算時間が 10000 秒を超えた場合は分枝限定法の計算を打ち切っているため，その場合は得られた変数集合および AIC 値は最適解とは限らない．

計算結果を分析する．まず，二種類のステップワイズ法によって得られた解は，選択された変数集合がかなり異なっていることがわかる．$\mathrm{SWR}_{\mathrm{const}}$ と $\mathrm{SWR}_{\mathrm{all}}$ では k の値がかなり異なるほか，特に `Automobile` では AIC 値が 20 以上異なり，これは無視できない差と言える．

次に，$p < 30$ であるような小さなデータセットに対しては，MISOCP は高速に最適解を計算できていることがわかる．また，大きなデータセットに対しては，10000 秒以内で最適解を求めることはできていないが，その場合でも二種類のステップワイズ法とほぼ同等またはそれ以上の解を出力できている．特に `Automobile` では，$\mathrm{SWR}_{\mathrm{const}}$ と $\mathrm{SWR}_{\mathrm{all}}$ の良い方と比較しても AIC 値が 10 以上小さな変数集合を得ることができている．

一方で計算時間に関しては，MISOCP はステップワイズ法よりかなり長い時間を要している．これは，ヒューリスティクスと厳密解法との違いではあるが，特に MISOCP では，分枝限定法の実行中に解くべき連続緩和問題が SOCP になり内点法が必要となるため，通常の線形整数計画問題と異なり双対単体法を利用したホットスタートが働かない．これが計算時間増大の理由であると考えられる．

5 まとめ

本稿では，AIC 最小化問題の混合二次錐整数計画問題 (MISOCP) による定式化を提案した．また，提案した定式化により計算機実験を実施し，ステップワイズ変数選択法との比較を行った．小規模なデータセットに対しては，提案手法は高速に最適解を求めることができ，大規模なデータに対してもステップワイズ法と同等またはそれ以上の解を求めることが可能なことを確認した．

参考文献

[1] H. Akaike, "A New Look at the Statistical Model Identification," *IEEE Transactions on Automatic Control*, Vol. 19, No. 6, pp. 716–723 (1974).

[2] F. Alizadeh and D. Goldfarb, "Second-Order Cone Programming," *Mathematical Programming*, Vol. 95, No. 1, pp. 3–51 (2003).

[3] K. Bache and M. Lichman: UCI Machine Learning Repository. University of California, School of Information and Computer Science, Irvine, 2013. http://archive.ics.uci.edu/ml

[4] E.M.L. Beale, "Two Transportation Problems," *Proceedings of the 3rd International Conference on Operational Research*, pp. 780–788 (1963).

[5] E.M.L. Beale and J.A. Tomlin, "Special Facilities in a General Mathematical Programming System for Non-Convex Problems Using Ordered Sets of Variables," *Proceedings of the 5th International Conference on Operational Research*, pp. 447–454 (1970).

[6] D. Bertsimas and R. Shioda, "Algorithm for Cardinality-Constrained Quadratic Optimization," *Computational Optimization and Applications*, Vol. 43, No. 1, pp. 1–22 (2009).

[7] E.J. Hannan and B.G. Quinn, "The Determination of the Order of an Autoregression," *Journal of the Royal Statistical Society*, Series B, Vol. 41, No. 2, pp. 190–195 (1979).

[8] IBM ILOG, IBM ILOG CPLEX 12.5, 2012.

[9] 小西貞則，北川源四郎：情報量規準．シリーズ・予測と発見の科学 2, 朝倉書店，東京，2004.

[10] M.S. Lobo, L. Vandenberghe, S. Boyd, and H. Lebret, "Applications of Second-Order Cone Programming," *Linear Algebra and its Applications*, Vol. 284, No. 1, pp. 193–228 (1998).

[11] The MathWorks Inc., MATLAB R2012b, 2012.

[12] R. Miyashiro and Y. Takano, "Mixed Integer Second-Order Cone Programming Formulations for Variable Selection," Technical Report, No. 2013-7, Department of Industrial Engineering and Management, Tokyo Institute of Technology, 2013. http://www.me.titech.ac.jp/technicalreport/h25/2013-7.pdf

[13] G. Schwarz, "Estimating the Dimension of a Model," *Annals of Statistics*, Vol. 6, No. 2, pp. 461–464 (1978).

[14] N. Sugiura, "Further Analysts of the Data by Akaike's Information Criterion and the Finite Corrections," *Communications in Statistics — Theory and Methods*, Vol. 7, No. 1 (1978), pp. 13–26.

[15] R.J. Wherry, "A New Formula for Predicting the Shrinkage of the Coefficient of Multiple Correlation," *The Annals of Mathematical Statistics*, Vol. 2, No. 4, pp. 440–457 (1931).